Course Syllabus: Introduction to Data Science

**TEXTBOOK:** Hands-on Machine Learning with Scikit-Learn, Keras, and TensorFlow. 2nd Edition, Aurelien Geron, 2019, O’reilly.

Textbook 2: Python for Data Analysis, 3rd Edition, Wes McKinney, O’reilly. (https://wesmckinney.com/book/)

**IMPORTANT DATES**:

Last Add Date January 17

Last Drop Date March 25

Final Exam May 3, 2024, 10:15 a.m. – 12:15 p.m.

**PURPOSE AND OBJECTIVE:** The course aims to equip students with fundamental data science skills, focusing on tools and concepts essential for acquiring, cleaning, managing, and analyzing datasets. Through practical applications, students will develop proficiency in narrative building and visual storytelling using data. The objectives include fostering critical thinking, promoting collaborative work, preparing for advanced studies and careers, and instilling ethical data practices. By the course's end, students will be well-versed in data-driven decision-making, poised for success in both academic and professional contexts.

**MATERIAL TO BE LEARNED:** Throughout the course, students will grasp foundational data science concepts, mastering techniques for acquiring, cleaning, and analyzing datasets using tools like Scikit-Learn, Keras, and TensorFlow. Practical skills will be honed through hands-on experience, emphasizing narrative building and visual storytelling. Critical thinking in data science, collaboration on projects, and ethical considerations will be integral components. The curriculum prepares students for advanced studies and careers while fostering a continuous learning mindset, encouraging exploration of additional resources and advanced topics beyond the standard curriculum.

**EVALUATION:** **GRADING SCALE:**

Homework 30% A 93-100% A- 90-92% B+ 87-89%

Projects 30% B 83-86% B- 80-82% C+ 77-79%

Midterm Exam 15% C 73-76% C- 70-72% D+ 67-69%

Final Exam 15% D 60-66% F < 60%

Class Participation 10%

**PREREQUISITES:** DT 230 "Introduction to Computer Programming", MT 131 "Statistics".

**ACADEMIC INTEGRITY AND ATTENDANCE POLICY:** Upholding academic integrity is paramount in this course, with severe consequences for violations. Plagiarism, cheating, and unauthorized collaboration can lead to failing grades for assignments or exams and referral for judicial review. Additionally, more than three unexcused absences will significantly impact your grade, and excessive absences may lower it naturally. It is essential to communicate in advance about any absence, providing a valid reason and documentation for excused absences. Quizzes and exams require students to show their work for full credit, emphasizing clarity in expressing calculator processes if used extensively. Cell phone use, including texting, is strictly prohibited. Familiarizing yourself with the current Student Handbook is crucial for understanding Academic Integrity policies, examination procedures, and the Attendance Policy, especially regarding excused absences. Adhering to these policies ensures a fair and enriching educational experience for all.

OFFICE HOURS:  MW 1:00 PM-2:00 PM, F 10:30 AM – 11:30 AM in room CSA 230. Other times can be arranged on request.

**COURSE DELIVERY:** Initially designed as a face-to-face, one-on-one course by appointment, this class will seamlessly transition to an online, synchronous format if pandemic-related circumstances necessitate. In the virtual setting, communication will be facilitated through Zoom, and all sessions will be recorded to accommodate students with limited internet connectivity. Quizzes and tests will shift to take-home formats, ensuring flexibility, and any presentations required will be conducted through recorded videos submitted to me. This adaptive approach aims to maintain the integrity of the learning experience regardless of the mode of delivery.

**Academic Misconduct**: See the current Student Handbook for the college's Academic Integrity policies as they pertain to examinations, plagiarism, classroom behavior, and the process for handling academic misconduct charges.

HOMEWORK POLICY

Homework assignments play a crucial role in reinforcing concepts learned in class and promoting individual understanding. In this course, the following policies govern homework submissions and assessments:

1. **Timely Submission:** All homework assignments are expected to be submitted by the specified deadline. Late submissions may result in a deduction of points, with the severity of the penalty increasing the longer the delay.
2. **Quality and Originality:** Homework solutions should reflect individual effort and understanding. Plagiarism or copying from external sources is strictly prohibited and will result in academic consequences.
3. **Clarity and Organization:** Clear presentation of solutions and organized work are essential. Use proper formatting, labeling, and explanations to ensure that your responses are easily understandable.
4. **Collaboration:** Unless explicitly stated otherwise, homework assignments are to be completed individually. Unauthorized collaboration may lead to academic penalties.
5. **Grading and Feedback:** Assignments will be graded based on correctness, completeness, and adherence to instructions. Constructive feedback will be provided to aid in your understanding and improvement.
6. **Resubmission:** In certain cases, resubmission of corrected assignments may be allowed after receiving feedback. However, this is at the discretion of the instructor and may be subject to specific guidelines.

**PROJECTS**

Projects are integral components of this course, providing opportunities to apply data science concepts to real-world scenarios. To ensure a successful and collaborative project experience, the following policies are in place:

1. **Team Formation:** Projects may be conducted individually or in teams, as specified by the instructor. Teams should be formed early, and collaborative dynamics should be maintained throughout the project duration.
2. **Project Proposal:** A project proposal outlining the objectives, scope, methodology, and expected outcomes must be submitted and approved by the instructor before the project commences. This ensures alignment with course objectives and provides guidance for successful project execution.
3. **Progress Updates:** Regular progress updates, including milestones achieved and challenges faced, are required. These updates foster transparency, allow for timely feedback, and ensure that projects are on track.
4. **Collaboration and Communication:** Effective communication within teams is crucial. Utilize designated channels for team communication, and promptly address any issues or concerns that may arise during the project.
5. **Documentation:** Thorough documentation of the project, including data sources, methodologies, and code, is essential. Clear documentation enables understanding, transparency, and reproducibility.
6. **Final Presentation:** Each team is required to deliver a final presentation showcasing the project's objectives, methodology, findings, and visualizations. This presentation provides an opportunity to communicate results effectively.
7. **Individual Contribution:** In team projects, individual contributions should be clearly delineated. Each team member is expected to actively participate and contribute to the project's success.
8. **Ethical Considerations:** Projects must adhere to ethical standards, respecting privacy, confidentiality, and legal requirements. Any ethical concerns should be addressed promptly.
9. **Submission Deadline:** Projects must be submitted by the specified deadline. Late submissions may be subject to grade deductions.

**CATALOG DESCRIPTION:**

A survey of major data science tools and concepts. Students will learn the process of acquiring, cleaning, managing, and analyzing data sets to produce insights and make data-driven decisions. They will also gain experience with narrative building and visual storytelling using data.

**SKILLS:** This course aims to cultivate a diverse set of skills essential for success in the field of data science. Students will acquire proficiency in data acquisition from various sources, master the art of data cleaning and preprocessing, and apply data analysis techniques using tools like Scikit-Learn, Keras, and TensorFlow. The curriculum emphasizes the development of narrative-building skills, enabling students to communicate data insights effectively. Additionally, students will gain expertise in visual storytelling through the creation of compelling visualizations using libraries such as matplotlib and seaborn. Critical thinking is fostered, encouraging a discerning approach to data science problems. Collaborative teamwork is emphasized through group projects, and ethical considerations in data practices are highlighted, including privacy and responsible data use. The course also instills project management skills, continuous learning mindset, and prepares students for advanced studies and careers in the dynamic realm of data science.

|  |  |  |  |
| --- | --- | --- | --- |
| Week | Dates | Topics Covered | Assignments Due |
| 1 | Jan 17-21 | Introduction to Data Science | - |
| 2 | Jan 24-28 | Built-In Data Structures, Functions, and Files | Exercise: Review basic Python data structures |
| 3 | Jan 31-Feb 4 | Built-In Data Structures, Functions, and Files | Assignment: File manipulation with Python |
| 4 | Feb 7-11 | NumPy Basics: Arrays and Vectorized Computation | Exercise: Array manipulation with NumPy |
| 5 | Feb 14-18 | NumPy Basics: Arrays and Vectorized Computation | Assignment: Solving mathematical problems with NumPy |
| 6 | Feb 21-25 | Getting Started with Pandas | Exercise: Data manipulation with pandas |
| 7 | Feb 28-Mar 3 | Getting Started with Pandas | Assignment: Exploratory data analysis with Pandas |
| 8 | Mar 6-10 | Data Loading, Storage, and File Formats | - |
| 9 | Mar 13-17 | **Spring Break - No Classes** | - |
| 10 | Mar 20-24 | Data Cleaning and Preparation | Exercise: Handling missing data |
| 11 | Mar 27-31 | Data Cleaning and Preparation | Assignment: Data transformation and cleaning |
| 12 | Apr 3-7 | Data Wrangling: Join, Combine, and Reshape | Exercise: Reshaping and pivoting data |
| 13 | Apr 10-14 | Data Wrangling: Join, Combine, and Reshape | Assignment: Merging datasets and reshaping |
| 14 | Apr 17-21 | Plotting and Visualization | Exercise: Data visualization with matplotlib |
| 15 | Apr 24-28 | Plotting and Visualization | Assignment: Creating informative visualizations |
| 16 | May 1 | Review and Conclusion | Final project presentation |